PDS Support for Structured Data Types

Response to RFI: Preparation for the Development of a Community-Based Roadmap for NASA's Planetary Data Services
Solicitation Number: NNH15ZDA012L
Submitted By: Robert Deen
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr, MS 168-300, Pasadena, CA 91109
Bob.Deen@jpl.nasa.gov
818-354-7492

Summary

The PDS needs to support structured data types, such as terrain meshes, Geographical Information System (GIS) data and XML data.

Rationale

The transition from PDS 3 to 4 solved a number of problems with out-of-control data formats that were not really suited to archive. PDS 3 was far too permissive in terms of what it allowed. However, PDS 4 has gone too far the other way, making it impractical or impossible to support certain kinds of data that are necessary for planetary data analysis.

A perfect example of what *not* to do is the Mastcam, MAHLI, MARDI (MMM) image data on MSL. The data supplied to PDS 3 is essentially the raw telemetry stream, with a supplied C program to decode the data into usable form. This is highly inconvenient for users, and the data is not usable without that program. PDS 4 would not allow this, instead insisting on an image raster with XML metadata. The raster is a much more usable, standard image format in which the data can be represented, and PDS 4 rightly demands this for archive purposes.

However, there are a number of kinds of data routinely generated and used by mission personnel that cannot practically be represented in the simple tables and arrays supported by PDS 4. While not an exhaustive list by any means, the tall tent poles from my perspective are listed below.

Terrain meshes

The Mars rover missions depend on stereo imagery, taking dozens of stereo pairs on an average day. Teams such as mine (MIPL/OPGS – Multimission Image Processing Lab/Operational Product Generation Subsystem) perform stereo analysis on these images, generating (among many other things) XYZ point clouds, which are delivered to PDS in image form.

What is *not* delivered though, at least in any useful form, are the terrain meshes. These result from analysis of the XYZ point clouds, and consist of connected geometric triangles with texture (image) information for each triangle. Terrain meshes are required by 3D rendering programs – most direct renderings of XYZ point clouds are unsatisfactory for scientific use, as they contain too many holes and do not unambiguously define the actual surface.

Terrain meshes have been delivered previously for MER to PDS 3, but in the format used by operations. This format (SGI Performer) is an obscure, binary format, not supported by any current tools, and not documented well. It is not an appropriate archive format, but it was (barely) better than nothing. PDS 4 would (rightly) not accept this format.

Recent developments, however, have enabled conversion of these meshes to ASCII-based industry standard formats such as the OBJ format, which *could* be suitable for archiving. However, there is no way to support this data in PDS 4.

We (MIPL/OPGS) are constantly asked about providing terrain meshes to external scientists and other users (such as visualization experts and even game designers). While occasional agreements providing special access are made directly with the Project, in general we must decline such requests, as the mesh data is not officially available to the public.

This situation will only get worse in the future. Technology had reached a point in the last few years where interactive rendering of 3D meshes is possible not only on traditional computers, but also on smartphones, pads, and even within Web browsers. The demand for these terrain meshes will only increase. While meshes can be derived from data supplied to PDS, it requires some effort. This effort should not need not to be repeated by every data user, since the missions themselves have already done – and validated – the work.

GIS data

Another form of structured data is Geographical Information System (GIS) data. While maps themselves are easily represented in PDS 4, increasingly the maps alone do not tell the whole story. GIS systems provide all kinds of overlays, such other coregistered maps, vector overlays, and labels. Vector overlays and labels in particular are not supportable in PDS 4. Their value is apparent when used for things like rover traverse path, labeling observation targets, geological contacts, strata outlines, classifications of terrain or geological units, elevation contours, etc.

These kinds of features are common in planetary GIS systems, such as Google Mars/Moon, USGS's PIGWAD, ASU's JMARS, or Washington University's Orbital Data Explorer footprint service. It should not be the role of PDS to supplant these systems. However, these systems are not archival in nature – they last only as long as their funding agency continues to support them. Once that support ends, the tool is lost – but more importantly, so is the data. Without some way to preserve for posterity the structured data used by these systems, the work of the people who created them will be lost. Perhaps more importantly, the source data backing up results in published scientific papers will be lost, making reproducibility of results difficult at best.

While maps *per se* are the purview of USGS rather than PDS, the same principles apply to *in-situ* data that are not traditionally considered maps but still fit within a GIS framework. Significant amounts of orbital and *in-situ* mission data contain a geospatial component that is best recorded and archived in a GIS-structured format. This data often provides the basis for making maps and other higher-order products. Examples include rover traverse data, instrument and camera footprints, target locations, and image viewsheds.

XML data

This category is a bit harder to define. There is a lot of data that is contained in structured XML. Ironically (since PDS 4 labels are XML based), there is no way in PDS 4 (or even PDS 3) to support XML as the *data* portion of an archive.

To the extent the XML data is actually metadata, it could simply be transferred to the PDS 4 labels. But this is not always the case. An example of this is marked-up text. This cannot be represented in an existing PDS 4 format, other than as a simple text stream – in which case the XML markup has no meaning for PDS. Any kind of hierarchically organized data would also fall into this category.

Although I don't have a specific example, XML data is becoming more and more popular due to its power and flexibility, and PDS will be increasingly digging itself into a hole if it cannot support XML data files.

Telemetry Data

As discussed above in the MMM example, raw telemetry is not a suitable archive format. However, there is value in preserving the originally telemetered data. This should never be allowed as the primary data set (as in the MMM case), but it should be allowed, and even encouraged, to supply raw telemetry *in addition to* a proper archive format. As such, the documentation and labeling standards would be quite a bit looser, but that's okay since almost all users will use the archive data. As Mike Malin (Mastcam/MARDI PI) said (private communication), we never know if someone in the future might have better decompressing technology that could extract additional information from the raw telemetry stream. This is true, but we should also not make users jump through hoops in order to access the primary mission data. Providing the telemetry stream (when it makes sense for the instrument) as an ancillary file along with proper archive data achieves the best of both worlds – making it easy for users to access the data they need, while still allowing in-depth re-analysis if needed sometime in the future.

Suggestions

I recommend that PDS undertake a study to determine how structured data like these can be archived properly. It is important that we do not go down the "anything goes" route that PDS 3 ended up on – the data must really be suitable for archive. But we should also adhere to industry standards where appropriate.

For example, the OBJ format, at least the subset needed to represent terrain meshes, could be an appropriate archive format. It is ASCII only, meaning it can be read with simple scripts if necessary. It is an industry standard, directly usable by most 3D rendering software packages. And the format itself is fairly simple; a competent software engineer could reverse engineer it with little to no documentation – meaning even if industry moves on in the future and no longer supports it, the data will still be usable.

Those criteria might be a good starting point for considering what kinds of structured formats are acceptable for archive: ASCII-only, industry standard, and amenable to reverse engineering.

As a side note, ASCII files can get extremely large (some meshes are hundreds of megabytes in OBJ format), so PDS may want to consider allowing some limited forms of compression for these.

Impacts

Simply put, this kind of structured data simply must be supported. PDS is not living up to its charter if it does not. This data is being produced, and used, anyway, and it is the future of data. If PDS does not step up to the plate to provide methods of archiving it, PDS will be marginalized. People will find a way to distribute and use this data, but it will not be properly archived and is at significant risk of being lost over time. Effort will be duplicated in re-creating it, provenance is likely to be lost, and scientific results using it will become irreproducible.

Keep PDS relevant in the future by archiving this valuable data.

About the Author

Robert Deen is a Principal at the JPL Multimission Image Processing Lab, where he’s been since 1987. He is the lead developer for the ground-based image processing software used by all the recent Mars *in-situ* missions (MER, PHX, MSL, InSight, Mars 2020) for operations and science processing. This software does stereo processing, meshes, mosaics, radiometric correction, and higher-order products such as slope maps and arm reachability products. He is also on the operations team for all these missions. He is instrumental in PDS image label design for these missions, and wrote much of the technical content for the SIS (Software Interface Specification) documents describing the imaging data for them. He is the PDS data provider for the MSL PLACES rover localization database. Recently he has been involved in the PDS 4 transition effort for InSight, and occasionally contributes to PDS design discussions.